

University of Dundee

## Digital Peer Assessment in School Teacher Education and Development

Topping, Keith

*Published in:*  
Research Papers in Education

*DOI:*  
[10.1080/02671522.2021.1961301](https://doi.org/10.1080/02671522.2021.1961301)

*Publication date:*  
2021

*Licence:*  
CC BY-NC-ND

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*

Topping, K. (2021). Digital Peer Assessment in School Teacher Education and Development: A Systematic Review. *Research Papers in Education*. <https://doi.org/10.1080/02671522.2021.1961301>

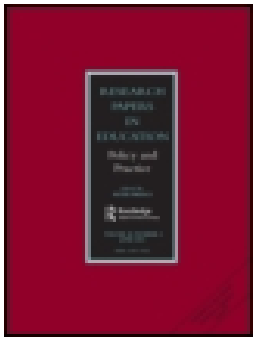
### General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



## Digital peer assessment in school teacher education and development: a systematic review

Keith James Topping

To cite this article: Keith James Topping (2021): Digital peer assessment in school teacher education and development: a systematic review, Research Papers in Education, DOI: [10.1080/02671522.2021.1961301](https://doi.org/10.1080/02671522.2021.1961301)

To link to this article: <https://doi.org/10.1080/02671522.2021.1961301>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 15 Aug 2021.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)

# Digital peer assessment in school teacher education and development: a systematic review

Keith James Topping 

University of Dundee, Dundee, UK

## ABSTRACT

Peer assessment (PA) is generally effective, and especially important for school teachers, as the experience might lead teachers to use PA more skilfully with school students. Digital PA (using computers) becomes more important as universities switch to online learning. This systematic review of research literature on digital PA for pre-service and in-service teachers encompasses: online/web-based, video, Massive Open Online Courses, digital frameworks to organize/structure PA, e-portfolios, Personal Digital Assistants (PDAs), Facebook, iPads and wikis. It contained 43 papers and all but one reported mainly positive effects. Potential moderator variables were considered, but studies rarely reported many of them. Few studies had control groups, only two reported Effect Sizes, and none reported implementation fidelity or follow-up. There was little evidence for carry-over of PA practices into later teaching of school children. However, the potential moderator variables provide a template for future reviews of PA and the design of PA by teachers.

## ARTICLE HISTORY

Received 16 March 2021

Accepted 20 July 2021

## KEYWORDS

Peer assessment; digital; school teachers; pre-service; in-service

## 1. Introduction

### 1.1. Purpose of this paper

Assessment is important for practising teachers, since as a major part of their job they need to undertake it effectively with diverse students. The research literature on teachers and assessment is almost all about the reliability and validity of assessment as practiced by teachers on students, largely focuses on summative assessment, and reports substantial variation and considerable error and bias in teacher assessments, even when for high stakes purposes (e.g. EPPI-Centre 2004; Harlen 2005). Teachers assess more reliably and validly when they participate in developing assessment criteria and/or use highly specific assessment rubrics.

This has led to concern about the reliability and validity of assessments made of teachers, Pecheone and Chung (2006) seeking to improve this by describing a structured assessment proforma (PACT) for assessing the performance of serving teachers. Darling-Hammond (2010) then broadened this debate, and Darling-Hammond, Newton, and Wei (2013) reported that PACT was highly predictive with pre-service teachers in terms

of teacher effects on student performance. However, such instruments do not appear to be widely used.

Assessment in teacher training proceeds by assessment of writing (especially of teaching programmes), presentations, self-reflection and observation of teaching practice – all largely assessed by professional staff. Assessment of a fully qualified and employed teacher is largely by writing of teaching programmes and observation of teaching practice, mostly done by head teachers or senior teachers, or by school inspectors. The emphasis is very much on summative assessment, and formative assessment is given little weight.

In this paper peer assessment is proposed as a vehicle for formative assessment for the performance of both pre-service and in-service teachers. The research evidence on its effectiveness in these contexts is systematically analysed, particularly where the peer assessment has taken place digitally. Given the importance of a deep understanding of assessment to teachers, might we expect that teachers would be better at peer assessment than other professions? In particular, might we expect that given a (hopefully positive) experience of peer assessment within their own professional community, teachers might be motivated and sufficiently skilful to deploy peer assessment with their pupils in class?

There is a gap in the research literature here, and the present paper offers the first systematic review of outcomes of PA conducted through digital technology in pre-service and in-service school teacher education and development, via online/web-based, video, Massive Open Online Courses (MOOCs), digital frameworks to organize/structure PA, e-portfolios, Personal Digital Assistants, Facebook, iPads and wikis. Results will be related to potential moderator variables, so that readers can obtain a richer picture of what is and is not present in the research.

### **1.2. *Why is peer assessment important?***

Feedback is widely considered important in education (e.g. Brooks et al. 2019) and PA is one method of enhancing the speed and quantity of feedback, if not the quality. Many professions may expect to engage in PA as part of their vocations. However, many teachers value PA but feel insecure about actually doing it (Deneen et al. 2019). Digital PA is likely to lead to further uncertainty, as many teachers are not yet comfortable with online professional development (e.g. Parsons et al. 2019). So, while the hope is that school teachers experiencing PA themselves can thereby be motivated and informed to undertake it effectively with students in the classroom, it is acknowledged that there are some challenges.

### **1.3. *Definition and types of PA***

A widely quoted definition of PA is: ‘an arrangement for learners to consider and specify the level, value or quality of a product or performance of other equal-status learners’ (O’Donnell and Topping 1998, 256). However, other similar terms (synonyms) are in the literature (e.g. peer grading/marking – giving a score to a peer product/performance; peer feedback – peers giving elaborated feedback; peer evaluation – more usually in workplaces regarding skill and knowledge; or peer review – more usually in academia regarding assessment of written papers).

Several previous studies compare two or three types of PA, but the variety in types of PA goes far beyond that. O'Donnell and Topping (1998) described a typology of relevant variables. Gielen, Dochy, and Onghena (2011) offered a more developed inventory. Further developments came from Topping (2018, 12–13), outlining 43 variables in the context of a comprehensive theory of PA. Many of these are used as potential moderator variables in relation to the independent variable of outcome in what follows.

## 2. Literature review

### 2.1. *Why might PA work?*

PA is not just for managing assessment burdens for teachers, but more importantly a mechanism for learning, particularly with elaborated feedback. For the assessor, the intellectual demands of reflecting, making a balanced assessment, formulating and delivering feedback can all lead to learning gains (Yu 2011). For the assessee, the intellectual demands of receiving and evaluating the feedback, deciding what aspects to implement and what not, and reflecting on other issues prompted by the feedback but not contained within it can all lead to learning gains (Li, Liu, and Zhou 2012). These mechanisms have been demonstrated by research over many years (e.g. Annis 1983). However, students new to PA might find it challenging and be particularly worried about a lack of respect shown by some assessors (Zhou, Zheng, and Tai 2020).

### 2.2. *Theoretical issues*

PA theory is rather scarce. In 2008 Friedman, Cox, and Maher used expectancy theory regarding students' motivation for PA, emphasising the belief that performance would lead to valued outcomes. In 2016 Reinholz advanced a model describing how peer assessment operated in marking/grading, analysis, feedback, conferencing and revision, noting that investigating learning opportunities was more useful than investigating student/instructor grade relationships.

A more comprehensive and integrated theoretical model was proposed by Topping (2018, chapter 4, 103–109), encompassing: organisation and engagement; cognitive conflict; scaffolding and error management; communication; affect; re-tuning and inter-subjectivity; practice and generalisation; reinforcement; metacognition, self-regulation and self-efficacy; and levels of learning. This indicated many of the processes which may or may not occur during PA.

### 2.3. *Social aspects of PA*

As university teaching is increasingly delivered online, and PA on a large scale is managed more easily online, interest in digital PA has grown. However, the social context of online learning is very different, especially when participants have no previous face-to-face experience. Van Popta et al. (2017) argued that while PA cognitive processes may be somewhat similar online and offline, social processes may differ. McLuckie and Topping (2004) compared the social, organisational and cognitive characteristics of effective peer learning interactions in face-to-face and online environments. In online PA, Cheng and

Tsai (2012) found that higher psychological safety, lower value diversity for goals, more trust in the self as an assessor and more positive social interdependence yielded deeper approaches to learning.

One feature of online PA is the affordance for anonymity, which has both advantages and disadvantages (Li 2017). Anonymity may be more important as PA starts, when social insecurity is at its height, but later may be less desirable. Further, online it may be easier to build in more consistent methods of scaffolding (Hou et al. 2020), although whether students use these is another issue. Cultural differences regarding acceptability of PA are another problem (Yu and Lee 2016). Students in countries where the predominant form of education is teacher-directed and not encouraging of independent thought may not like PA.

#### ***2.4. Previous outcome research on PA in general***

Irrespective, does PA work? The evidence on PA with all kinds of learners is generally positive, from the earliest reviews (e.g. Topping 1998, on peer grades and feedback; Falchikov and Goldfinch 2000, on peer grades) to the latest meta-analyses (e.g. Li et al. 2020; Double, McGrane, and Hopfenbeck 2020). Li et al. (2020) found an overall Effect Size (ES) of 0.29 in 58 studies (ESs are a quantitative measure of the magnitude of an effect), with significant moderator variables of Training & Online/Digital (moderator variables are third order variables that affect the size or nature of the relationship between an independent and dependent variable). Double, McGrane, and Hopfenbeck (2020) found an overall ES of 0.31 in 54 studies, but no significant moderator variables.

In PA studies, it is often assumed that teacher ‘expert’ assessment should be the criterion for validity, but both these studies showed PA was more reliable and had higher ESs than teacher assessment, although teacher assessment is not very reliable (e.g. Johnson 2013). There were similar effects at primary school, secondary school and in higher education.

#### ***2.5. Previous outcome research on digital PA***

It is unsurprising that online PA has been separately reviewed (e.g. by Tenorio et al. 2016; Fu, Lin, and Hwang 2019). The first meta-analysis, by Zheng, Zhang, and Cui (2020), found 37 controlled studies 1999–2018. Eight studies were in school and the rest in higher education, and this mixing of contexts is a weakness. Of the 37 studies, 19 examined outcomes (overall ES 0.58 – moderate) and 17 the effects of extra supporting strategies (ES 0.54).

These ESs are larger than those reported most recently for PA in general (above), suggesting that (despite some disadvantages), digital PA has countervailing advantages that make it more effective. Training and anonymity improved outcomes, and duration of PA was also important (6–10 weeks being the optimum). However, direct comparison of online and offline learning was rare – most studies compared online PA to no PA. None of the papers in Zheng, Zhang, and Cui (2020) referred to teachers.

## 2.6. Nomenclature

Here, ‘pre-service teachers’ means teachers in college/university prior to commencing teaching; ‘in-service teachers’ means teachers already in teaching posts (whether qualified or not) and receiving further training. Both may be referred to as ‘students’. ‘Students in schools’ means children in a school classroom. ‘Teachers’ means teacher educators or others with responsibility for managing/assessing student work. Likewise, the terms pre-service education/training equate with Initial Teacher Education (ITE), while in-service education/training equates to Continuous Professional Development (CPD).

## 3. Research questions

Developing in particular from [sections 1.3](#) and [2.5](#) above, two research questions were posed:

- (1) Does digital PA in school teacher education and development have more positive effects than negative effects (minimally in the ratio 2:1 per study)?
- (2) Do all studies refer to a range of potential moderator variables, and if not, what gaps are evident?

## 4. Methodology for systematic review

Alexander (2020) offered an analysis for education professionals of the PRISMA systematic review and meta-analysis principles (which were principally designed for medical professionals). Nevertheless, here we report under the conventional PRISMA categories.

### 4.1. Timespan

Items on digital PA were not expected before 2000, so the search was limited to 2000–2020. Occasionally items turned up which were dated before 2000, however, and were included if they met the inclusion criteria.

### 4.2. Language

Only papers in the English language were included.

### 4.3. Databases

Databases searched were: ERIC, Scopus, Science Direct, APA PsychNet, JSTOR, the Social Science Research Network, CORE, ResearchGate, the Directory of Open Access Journals, EBSCO Teacher Reference Centre, JURN, Semantic Scholar, Paperity and Google Scholar. Given the breadth of these databases, referential backtracking was not undertaken.

#### **4.4. Search terms**

Initial Search Terms for each database were: 'peer assessment' AND technology AND 'teacher education' OR 'teacher training'. The search was then re-run in each database using the terms: 'peer grading' OR 'peer marking' OR 'peer review' OR 'peer evaluation' OR 'peer feedback' instead of 'peer assessment'.

#### **4.5. Publication status**

Journal articles, conference papers ( $n = 6$ ), research reports ( $n = 2$ ) and Ph.D. theses ( $n = 1$ ) were accepted. Books and book chapters were not accepted, as usually they had not had any kind of peer or external review.

#### **4.6. Inclusion criteria**

The item had to: (1) include participants who were primary, middle or secondary school teachers or in education or training to be such, (2) investigate PA (or synonyms thereof) in the context of some form of digital technology; (3) give relevant empirical data (quantitative or qualitative or both) on outcomes or effects; (4) be in the English language; (5) be published in 2000 or later. Where qualitative studies were included, an indication of number of participants holding any view or opinion on outcomes or effects was required, and if this was not present, the study was rejected. Two-thirds of participants needed to hold any given opinion (whether positive or negative) for the study to be included.

#### **4.7. Screening, eligibility, selection**

The abstracts of 661 eligible papers (excluding duplicates) (and full papers where necessary) were carefully screened by two senior researchers both very familiar with the area, and decisions made about selection for inclusion. Inter-rater reliability was 0.92, satisfactorily high. All disputed studies were resolved by discussion. Thirty-four papers on PA were selected and the synonyms yielded 9 (Figure 1 gives the PRISMA data; Moher et al. 2009). The full text of these 43 papers was then carefully read.

#### **4.8. Measures**

A majority of studies reported outcome differences on relatively objective measures in mean difference between intervention and control groups or pre- and post-testing, but a minority of studies reported only on relatively subjective measures, usually with only one occasion of data-gathering. This is further explained below.

#### **4.9. Quality of evidence**

Quality of Evidence was initially categorised from 1 to 4 using the GRADE framework of Guyatt et al. (2011), but inter-rater reliability was too low and this method was eventually considered too subjective and unreliable for weighting purposes.



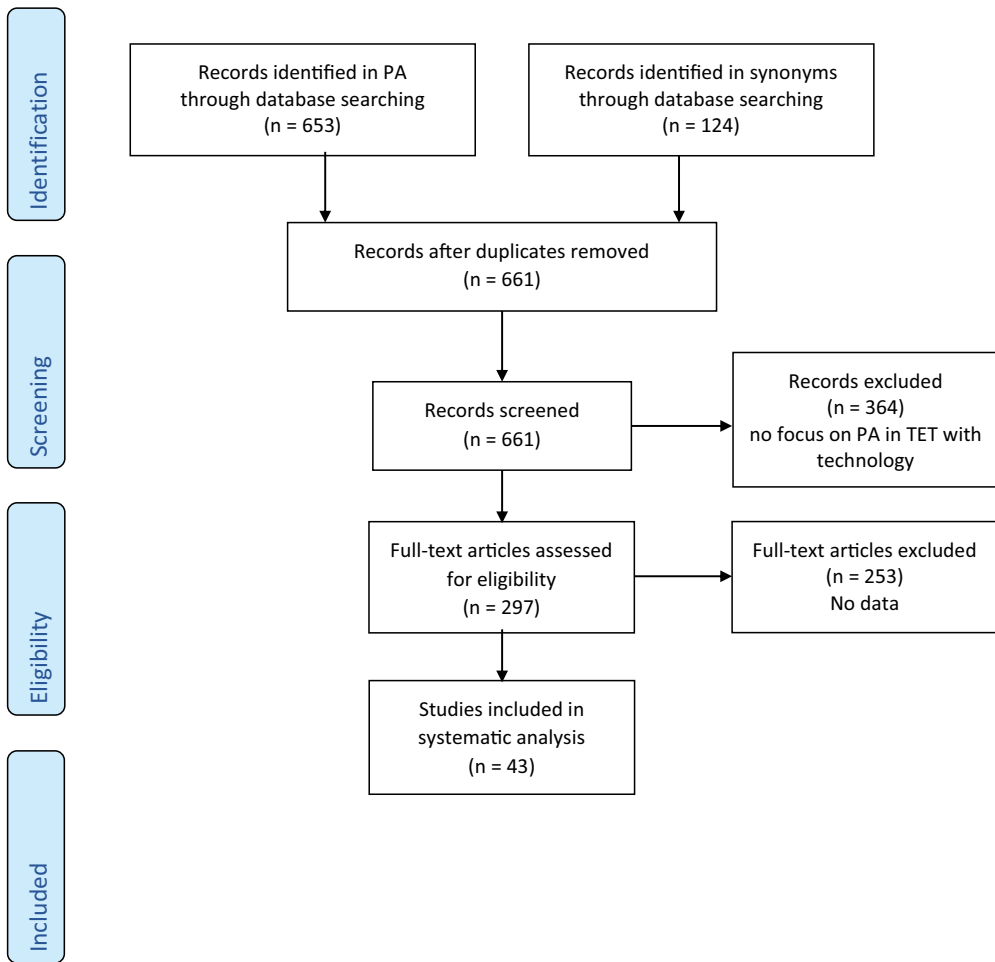


Figure 1. PRISMA flow diagram.

#### 4.10. Potential moderator variables

Unlike meta-analyses, where potential moderator variables can be assessed for statistical significance in relation to ESs, a systematic review enables the investigation of potential moderator variables which are theoretically relevant. In this study, the potential moderator variables were drawn from [sections 1.3](#) Typology and [2.2](#) Theory above. Readers can pursue the references given there if they wish for further definition of the variables or an idea of how they might cluster together.

The main variables chosen were: Study Authors/Date, Type of Technology, Country of Origin, Pre-service or In-service Education/Training, Number of Participants, Gender Balance, Subject Studied, Year of Degree, Length of Intervention, Grades or Elaborated Feedback or Both, nature of the Product Assessed and the nature of the Measure(s).

The secondary variables chosen were: Compulsory vs. Voluntary, Cognitive and/or Metacognitive gains, Social or Emotional goals, Motivation/Self-efficacy, previous Familiarity with PA, Pairs or Groups, Matching Strategy, Consistency of Matching,

Frequency, Duration, Anonymous or Identified, Participants Design Criteria or not, provision of Training, Synchronous or Asynchronous, Location, Time, use of Rubric, additional Prompts or Scaffolds, guidance given on Form of Feedback, guidance given on Justification of Opinions, Complexity of Work, Individual Contribution to group effort, Feedback by Discussion or Writing, Formative or Summative, to include suggestions for Improvement, Supplementary or Substitutional, Curriculum Alignment, alignment with subsequent Formal Assessment, Monitoring or Quality Control, contribution to Final Formal Assessment, effect of Cultural Variables, other intrinsic or extrinsic Rewards, Implementation Integrity, Follow-Up, presence of ESs and presence of Theory.

#### **4.11. Coding**

For consistency, the 12 main and 37 secondary moderator variables were then coded by a single senior researcher, who appraised the extent to which studies had considered/ reported the 49 moderator variables. Test-retest reliability was very high (0.98). This discrimination of many moderator variables was likely to be helpful for researchers who might otherwise blindly combine very various methods, and for teachers who might wish to use the variables to guide the planning of new PA projects.

### **5. Results**

#### **5.1. Overall outcome**

All but one of the 43 studies reported mainly positive effects (more than two-thirds positive), although some noted some initial student resistance. The only negative finding was in one of nine studies of video (Picci, Calvani, and Bonaiuti 2012). Table 1 outlines findings for the main variables. All moderator variables are then discussed in three sections below: those present with variability, those present but largely similar, and those absent or largely absent.

#### **5.2. Moderator variables present with variability**

##### **5.2.1. Type of technology**

Previous studies of PA have regarded online/offline as a variable, but the varieties of digital PA are much wider than that:

**5.2.1.1. Online and web-based.** The twelve papers here were very various, especially in terms of what was peer assessed. Seven focused on PA of specific materials: three on proposals (Chen and Tsai 2009; Tsai 2012; Wen and Tsai 2008), two on teaching materials (Sert and Aşık 2020; Tsai and Liang 2009), one on projects (Tsivitanidou and Constantinou 2016) and one on written assignments (Seifert and Feliks 2019). The remainder dealt with web-based case conferencing (Bonk et al. 2001), weblogs (Wopereis, Sloep, and Poortman 2010), the effect of training (Liu and Li 2014) and PA over time (Tsai, Lin, and Yuan 2002; Tsai et al. 2001).

**Table 1.** Principal moderator variables in systematic review of digital peer assessment in school teacher education and development (N = 44).

Authors	Type of technology	Country of origin	Pre-service or In-service (P)	N	Gender (F = female, M = male)	Subject, Year of degree	Length of intervention (months)	Grades (G) or Feedback (F) or Both (B)	Product	Outcome data
Wen and Tsai (2008)	Online: Proposals	Taiwan	I	37		Science technology	2	B	Research methods proposal	Peer scores
Chen and Tsai (2009)	Online: Proposals	Taiwan	I	52	22 F 30 M	Science technology	1.5	B	Research methods proposal	Peer scores
Tsai (2012)	Online: Proposals	Taiwan	I	40		Science technology	2		Research methods proposal	Interview
Tsai and Liang (2009)	Online: Teaching Materials	Taiwan	P	36		Pre-school education	2	B	Science project for preschool children	Peer scores
Sert and Aşık (2020)	Online: Teaching Materials	Turkey	P	111		EFL	12	F	Computer Aided Language Learning Digital learning products	Keyword analyses of blogs Pre- post test
Tsitivanidou and Constantinou (2016)	Online: Projects	Greece	P	27		Science, 4 <sup>th</sup> year		F	Assessment performance	Survey
Seifert and Feliks (2019)	Online: Written Assignments	Israel	P	300				F	Case-based discussion	Peer feedback
Bonk et al. (2001)	Online	USA	P	700		Educational psychology 1 <sup>st</sup> year + 3 <sup>rd</sup> year	24	F	Reflection weblogs	Weblog content
Wopereis, Sloep, and Poortman (2010)	Online	Netherlands	P	20	8 F 12 M	Early childhood	2	F	Own Project	Peer scores
Liu and Li (2014)	Online	USA	P	78	63 F 15 M	Science		B	Designing science activities	Work quality
Tsai et al. (2001)	Online	Taiwan	P	24		Science	2	B	Designing science activities	Peer scores
Tsai, Lin, and Yuan (2002)	Online	Taiwan	P	24	11 F 13 M	Physical education		F	Lesson delivery	Surveys, interviews
Lamb (2012)	Video: Peer Feedback	UK	P	23	6 F 17 M					

(Continued)

Table 1. (Continued).

Authors	Type of technology	Country of origin	Pre-service (P) or In-service (I)	N	Gender (F = female, M = male)	Subject, Year of degree	Length of intervention (months)	Grades (G) or Feedback (F) or Both (B)	Product		Outcome data
									Lesson delivery to peers	Survey	
Savas (2012)	Video: Peer Feedback	Turkey	P	40	34 F 6 M	EFL	4	F			Survey
So, Hung, and Yip (2008)	Video: Peer Feedback	Hong Kong	P	3		Pedagogy		B	Video of teaching practice	Quality of discussion	
Borowczak and Burrows (2016)	Videos: Peer Feedback	USA	P	27		STEM	24	F	Create 2 videos; post to Youtube	Video rating	
Wu and Kao (2008)	Video: Tagging	Taiwan	P	36		ICT	3	G	Micro-teaching or field teaching	Survey	
N Picci, Calvani, and Bonaiuti (2012)	Video: Tagging	Italy	I	13	13 F	Pedagogy	0.5	F		Pre-post tests	
Çelik, Baran, and Sert (2018)	Video, tagging, Mobiles	Turkey	I	2		EFL	1	F	3 teaching sessions	Quality of observation	
Goeze et al. (2014)	Video: Hyperlinks	Germany	P	100	77 F 23 M	EFL	0.13	F	Six teaching videos	Content Analysis	
Luo, Robinson, and Park (2014)	MOOC: Supports	USA		1825			1.25	G	Mapping technology	Peer scores, Survey	
Bachelet, Zongo, and Bourelle (2015)	MOOC: Supports	France	P + I	4650	5285 M		0.17	B	Tasks: Concept map, Gantt chart	Peer grades	
Gamage et al. (2017)	MOOC: Supports	Sri Lanka	P + I	87		Creativity and innovation		G	Grading open ended questions	Feedback quality score	
Laurillard (2016)	MOOC: Supports	UK	P + I	174		ICT	1.5	B	Case studies + tasks	Survey	
Vivian, Falkner, and Falkner (2014)	MOOC: Supports	Australia	I	174	F148 M26	Computer science			Creation of teaching resource + 2 lesson plans	Survey	
Mohamed and Hammond (2018)	MOOC: Supports	UK	P + I			EFL	4	G	Teaching techniques & strategy		
Sterbini (2013)	MOOC: Supports	Italy	P + I	136			PA of open answers	G		Peer grades	
De Marsico et al. (2018)	MOOC: Supports	Italy	P + I	1000				G		Peer grades	

(Continued)

Table 1. (Continued).

Authors	Type of technology	Country of origin	Pre-service (P) or In-service (I)	N	Gender (F = female, M = male)	Subject, Year of degree	Length of intervention (months)	Grades (G) or Feedback (F) or Both (B)	Product	Outcome data
Anat, Einav, and Shirley (2020)	MOOC: Supports	Israel	P	29		Mathematics 4 <sup>th</sup> year		F	Design collaborative learning environment	Work quality
Vu (2017)	MOOC	Vietnam	P + I	1215	642 F 573 M	EFL	1.6	B		Survey, interview
Gogoulou et al. (2007)	Digital Frameworks	Greece	P	35		Informatics	2	F	Project, topic given	Survey
Põldoja et al. (2012)	Digital Frameworks	Estonia	I	50			0.5	G		Profile building
Foschi and Cecchinato (2019)	Digital Frameworks	Italy	I	42			6	B		Peer scores
Chang (2002)	E-portfolios	Taiwan	P	35		Computers and Instruction 1 <sup>st</sup> year	1.5			Survey + Interview
Welsh (2012)	E-portfolios	UK	P	170	102 F 12 M		7			Survey
Chen (2010)	PDAs	Taiwan	P	37	25 F 12 M		3	G	Lesson plan; 2 projects	Survey
Okumuş and Yurdakal (2016)	Facebook: Sharing Video	Turkey	P	38	29 F 9 M	EFL	3.25	F	Microteaching	Survey
Lin (2016)	Facebook: Sharing Video	Taiwan	P	31 (E16 + C15)	20 F 11 M	1 <sup>st</sup> year	2.5		Microteaching	Survey
Ersöz and Sad (2018)	Facebook: Sharing Video	Turkey	P	10						Interviews
Demir (2018)	Facebook: Sharing Materials	Turkey	P	28	15 F 9 M	Science	1.5	F	Materials for teaching	Interviews

(Continued)

Table 1. (Continued).

Authors	Type of technology	Country of origin	Pre-service (P) or In-service (I)	N	Gender (F = female, M = male)	Subject, Year of degree	Length of intervention (months)	Grades (G) or Feedback (F) or Both (B)	Product	Outcome data
Ramdani and Wrdodo (2019)	Facebook: Sharing Materials	Indonesia	P	40		EFL	3.75			Oral language
Backhouse, Wilson, and Mackley (2015)	iPads	UK	P	140		Early years & Primary Education	8	B	Presentations	Feedback quality + observations
Ng (2016)	Wikis	Hong Kong	P	76	76 F	Early years ICT	3	B	Group project. Evaluate website for children	Project quality

Where a cell is blank, this information was not available in the paper. N = the only main negative outcome reported.

**5.2.1.2. Video.** Four of the nine papers dealt with peer feedback on real-life videos (Lamb, Lane, and Aldous 2012; Savas 2012; So, Hung, and Yip 2008; Wu and Kao 2008). Two dealt with tagging or annotating such videos (Çelik, Baran, and Sert 2018; Picci, Calvani, and Bonaiuti 2012, – the latter being the only negative finding). One concerned PA of YouTube videos (Borowczak and Burrows 2016) and another video and hyperlinks (Goeze et al. 2014).

**5.2.1.3. MOOCs.** Four reviews of Massive Open Online Course (MOOC) literature were found, but were not focused only on teachers (Bozkurt, Akgün-Özbek, and Zawacki-Richter 2017; Jacoby 2014; Kennedy 2014; Veletsianos and Shepherdson 2016). The ten papers in this section mainly described various supports for PA that had been built into MOOCs, arguably necessary given the very low completion rate (Bachelet, Zongo, and Bourelle 2015; De Marsico et al. 2018; Gamage et al. 2017; Laurillard 2016; Luo, Robinson, and Park 2014; Mohamed and Hammond 2018; Sterbini and Temperini 2013; Vivian, Falkner, and Falkner 2014). Anat, Einav, and Shirley (2020) investigated student *creation* of parts of a MOOC. There was one other paper (Vu 2017).

**5.2.1.4. Digital software to organize/structure PA.** Technological tools are available to organise, structure and support PA (Luxton-Reilly 2009) (e.g. Expertiza <http://wiki.expertiza.ncsu.edu>; PeerScholar <https://doc.peerscholar.com>), but few specifically target teachers. However, three papers did (Foschi and Cecchinato 2019, <https://www.peergrade.io>; Gogoulou et al. 2007, <http://hermes.di.uoa.gr:8080/scale>; Pöldoja et al. 2012).

**5.2.1.5. E-portfolios.** Two positive papers explored this area (Chang 2002; Welsh 2012).

**5.2.1.6. Personal digital assistants (PDAs).** One paper investigated the successful use of PDAs (Chen 2010).

**5.2.1.7. Facebook.** Three studies used Facebook largely as an effective platform for sharing and PA of videos or pictures (Ersöz and Sad 2018; Lin 2016; Okumuş and Yurdakal 2016). Another two investigated Facebook in PA of materials and language (Demir 2018; Ramdani and Widodo 2019).

**5.2.1.8. iPads.** One study investigated the use of iPads, which increased engagement and critical reflection (Backhouse, Wilson, and Mackley 2015).

**5.2.1.9. Wikis.** One study investigated the creation of Wiki sites for school children (Ng 2016).

## 5.2.2. Country of origin

For the MOOCs (where participants can enrol from all over the world), the creating country was Italy (2), the UK (2), the USA (2), Australia, France, Israel and Sri Lanka. For other technologies, participants came from Taiwan (10), Turkey (6), the UK (3), the USA (3), Greece (2), Hong Kong (2), Italy (2), Israel, Germany, the Netherlands, Estonia and Indonesia. Taiwan is clearly over-represented.

### 5.2.3 *Nature of participants*

In 27 studies the participants were pre-service teachers, in eight studies in-service teachers, and in eight studies (largely MOOCs) they could be either pre-service or in-service teachers. MOOCs might include past, present or future teachers, but these studies were unspecific about the vocational composition of the enrolled or completing population.

In 42 studies there were 11,715 participants, ranging from 2 to 4650 (MOOCs tended to have much higher numbers). In one study participant numbers were unclear. The mean was 279 but 29 studies had numbers of participants only in double figures. Regarding gender of participants, so far as this was reported there were 3556 females and 6053 males (although the very large number of males in Luo, Robinson, and Park 2014 contributed greatly). As most teachers are female, this finding was counter-intuitive.

### 5.2.4. *Nature of PA*

PA was particularly common in Science and English as a Foreign Language (EFL) (eight studies each). Information technology (ICT) and computer science were also popular (five studies each). Early Years education (three studies) and Pedagogy (two studies) were somewhat popular, together with single studies in Creativity, Educational Psychology, Informatics, Mathematics and Physical Education. Twelve studies were unclear. Of the pre-service studies, only five reported the academic year in which the PA occurred (neglecting the opportunities for follow-up in subsequent years).

The length of intervention was very variable, ranging from 4 days to 24 months, but 11 studies did not report this. The average was 4.15 months, but 18 studies had intervention lengths of 2 months or below. Feedback was more popular (15 studies) than Grades (eight studies), but 13 studies combined the two methods. Seven studies were unclear.

The products to be peer assessed were quite various, but ten involved the delivery of a lesson, either to peers in a micro-teaching setting or to real school students in a classroom. One further paper required the development of a lesson plan. Three papers required the development of teaching materials, and two teaching techniques. Four papers required the development of a project for school children, while a further two focused on the participant's own project. Three papers required the development of a research proposal. The remaining nine papers all had different products.

The nature of the outcome data was also very variable (here nature of outcomes has been counted rather than number of papers). There were 14 reports of survey data, 11 of peer grades or scores, and ten of work quality. Six reported on interviews, two on peer feedback and two on pre-post tests. Single papers reported on video rating and profile building respectively. Thus, 20 papers were based on subjective perceptions, while 25 offered more objective data.

## 5.3. *Moderator variables present but mostly similar*

Most studies characterised PA as formative (rather than summative) and expected it to include suggestions for improvement. Likewise, most studies (except for MOOCs) reported PA was supplementary rather than substitutional, although two studies reported it as substitutional (Luo, Robinson, and Park 2014; Bachelet, Zongo, and Bourelle 2015). Thus, in many studies, teachers were still also assessing all products themselves. Alignment



with the curriculum was generally reported as high. However, alignment with any subsequent summative form of assessment was not reported (except for Gamage et al. 2017).

Involving participants in designing assessment criteria enhances their engagement with PA, yet only two papers reported doing this (Liu and Li 2014; Lamb 2012). Providing training enhances outcomes from PA, but only 13 (30%) studies reported this. The amount of training varied greatly, more extensive training being reported by Bachelet, Zongo, and Bourelle (2015), Goeze et al. (2014) and Welsh (2012). In four cases the training also involved practice at PA.

Half of the studies ( $n = 22$ ) used a rubric to scaffold the PA, although in some cases this was very simple. However, only two studies reported deploying additional prompts or scaffolds, via hyperlinks or case studies. The complexity of work assessed was generally high (reported in 13 studies), but there was no evidence of PA occurring with simpler tasks before progressing to harder tasks. It is sometimes useful to evaluate the contribution of each participant to the group effort, and five studies reported the quality of feedback given by individuals was also subject to PA, which is a version of this. Ten studies reported quality control by comparing PA to instructor assessment, although instructor assessment is less reliable than PA (Johnson 2013). No other quality control was reported.

#### **5.4. Moderator variables showing variability**

Participation in PA in MOOCS was voluntary and some participants chose not to do it. Beyond this, nine studies had PA as compulsory, six as voluntary, and 19 did not specify. Sixteen studies targeted cognitive gains, while nine studies targeted cognitive and meta-cognitive gains – others were unspecific. No study mentioned social/emotional goals or motivation/self-efficacy.

PA was most usually arranged by dividing the participants into groups of three (occasionally four) and having them all assess each other, perhaps because the average was felt to be more reliable, but this was rarely stated. A few studies grouped 6–8 participants, but here the amount of PA work might be perceived as excessive. Four studies had participants in reciprocal pairs, perhaps reducing reliability but increasing the likelihood of social bonding. One study had groups of three producing a joint product which was peer assessed by another group. The rest were unspecific.

Matching seemed to have been given little thought. Many studies reported same-age matching and no study reported cross-age matching. Some studies reported random matching, but otherwise it appeared chaotic. Only one study reported matching by ability, grouping participants for PA according to GPA (Demir 2018), although others matched according to interest (Chen and Tsai 2009) or assigned an assessor based on the quality of their previous feedback (Gamage et al. 2017). Regarding consistency of matching over successive rounds of PA, six (out of 43) reported deliberate consistency, while one reported deliberately using different assessors for each round.

Regarding the frequency of PA (and consequently the extent to which participants were likely to become used to it), studies mostly reported three rounds of PA ( $n = 6$ ). Single studies reported one, two, five or eight rounds respectively, extending over 4–9 weeks. It is surprising more studies did not report this, as it seems likely to affect acceptability and possibly effectiveness. Studies were equally divided on anonymous or

identified PA (six studies each; the remainder unspecific). Of the identifying studies, three reported deliberate public availability of PA outcomes. Giving feedback by discussion seemed more popular than giving feedback in writing.

### **5.5. Moderator variables absent or almost absent**

Two papers noted participants were familiar with PA, three noted participants were not, and 38 made no comment. Given that initial student resistance to PA is commonly reported, this omission is surprising. Only one paper reported on the time or place when engagement in PA occurred, and this was in class. In other papers the engagement might have been out of class at any time convenient to the interacting participants, although this was not stated, nor were any difficulties in arranging meetings mentioned, nor were the interactions characterised as synchronous or asynchronous.

No paper reported on guidance on form of feedback – whether it should be positive or negative or both and in what quantity. Only one paper (Tsai, Lin, and Yuan 2002) gave any details of the kind of reflection or justification of opinions expected. Only one study reported that PA contributed to the final assessed outcome (Luo, Robinson, and Park 2014), and this was in a MOOC. One might assume that this would be true for other MOOCs, but it was not stated. Three studies specifically said that there was no contribution of PA to final assessed outcomes. There was no reporting of any other rewards.

Even though these studies came from all over the world, no study commented on the effect of cultural variables on PA. No paper made any mention of implementation fidelity/integrity or follow-up. The five studies including PA of quality of peer feedback did have a kind of measure of implementation fidelity/integrity, but this was never argued. Only two papers gave ESs. Although peer assessment could be theorised in terms of various socio-cognitive and other models, none of this featured in the papers scrutinised, although one paper did mention Activity Theory (Gogoulou et al. 2007).

### **5.6. Prospects for future meta-analysis**

Only 21/43 papers yielded quantitative outcome data amenable to generating ESs, and only three had control or comparison groups. Only two papers reported ESs, so otherwise these were calculated. A total of 4422 participants in 21 projects generated 105 ESs over an average period of 2.5 months. None were negative and the mean ES was 0.57 (s.d. = 0.40; 95% Confidence Interval 0.49–0.64). The Control/Comparison Group studies (n = 3) yielded a higher overall ES of 0.75 (s.d. = 0.51; 95% Confidence Interval 0.27–1.22), but their underlying distribution was not normal. The other studies reported either Surveys (subjective perceptions) (n = 8) or Peer Scores on various PA instruments (n = 11), in both cases with underlying normal distributions. However, given that the number of control group studies was so small and that weak measures were often used, the field is not yet ready for meta-analysis. Here none of the moderator variables reached significance, although some differences were quite large.

### 5.7. Publication bias

There was an insignificant negative correlation ( $r = -0.26$ ) between study sample size and mean ES. Orwin's (1983) failsafe N indicated that 288 missing studies would be needed to bring the ESs under the .05 level. This was far higher than the existing 21 and suggested that unpublished data was unlikely to have influenced the results.

## 6. Discussion

### 6.1. Summary

The search uncovered 43 papers meeting the inclusion criteria for the systematic review. Twelve of these were on online/web-based PA, nine used video, ten were on MOOCs, three were on a digital framework for PA, two used e-portfolios, one used PDAs, five used Facebook, one used iPads and one focused on Wikis. Twenty-seven papers focused on pre-service teachers, eight on in-service teachers, and eight on both. All but one (Picci, Calvani, and Bonaiuti 2012) had positive main findings, although some also had a minor negative finding (particularly regarding time and effort consumption, consistency and objectivity). Studies were related to 49 moderator variables, although no study reported on all of them. Different sections reported on Moderator Variables Present with Variability, Moderator Variables Present and Mostly Similar, and Moderator Variables Absent or Almost Absent.

### 6.2. Critique of methodology of studies reviewed

The quality of studies overall was generally weak, with issues pertaining to design, sampling, measures and analysis. Sample sizes were very various (2–1825). Most studies were of short-term interventions, and the lack of any follow-up was concerning. Few of the studies used random allocation to conditions, and were not analysed with techniques appropriate for clustered data (e.g. multi-level modelling), so teacher effects may have had considerable impact. In many studies, peers were both assessors and assessees, but no studies explicated the gains ensuing from being one or the other. Only two studies gave ESs and none mentioned implementation integrity or fidelity. Perhaps the most troublesome failing was the absence of control/comparison groups. The situation for PA in in-service teacher development is less known, as there were few studies. There was no evidence on subsequent successful use of PA with school children, although Yim and Cho (2016) tried to predict this. All of these should be remedied in future research.

However, the search for literature was extremely thorough, in 14 databases, and publication bias was small. PA in teacher education and development is clearly successful and widely spread. This suggests that PA could be useful in many countries.

### 6.3. Relationship to previous literature

In wider meta-analyses of PA, Double, McGrane, and Hopfenbeck (2020), Li et al. (2020), and Zheng, Zhang, and Cui (2020) found PA to be effective, so the positive finding in this systematic review of PA with teachers is unsurprising, although this study is the first with this focus. Double, McGrane, and Hopfenbeck (2020) and Li et al. (2020) meta-analysed

studies in both digital and analogue PA for all vocations, finding small to medium ESs. Zheng, Zhang, and Cui (2020) focused only on digital PA for all vocations, finding a medium overall ES. Li et al. (2020) noted that digital PA was significantly more effective than analogue PA, confirming the Zheng, Zhang, and Cui (2020) finding. While Double, McGrane, and Hopfenbeck (2020) found no significant moderator variables, both Li et al. (2020) and Zheng, Zhang, and Cui (2020) found training led to higher ESs than no training, as did the present study. In this study, duration of intervention was not a significant variable, but Zheng, Zhang, and Cui (2020) found that duration made a difference.

#### **6.4. Critical analysis of strengths and weaknesses**

Digital PA is likely to become more important than analogue PA as more universities shift to more online learning. For school teachers, their experience of PA in initial education or continuous professional development should prepare them for using it subsequently with students in schools, but there is very little evidence that this is happening at the moment, and studies are urgently needed which investigate any such extended effect.

Considering moderator variables, the main ones were: type of technology, country of origin, pre-service or in-service education/training, number of participants, gender balance, subject studied, year of degree for assessors and assesseees, length of intervention, grades or elaborated feedback or both, nature of the product assessed and the nature of the measure(s). These form a minimum reporting requirement for any study purporting to investigate this area.

There was little difference in outcomes according to type of technology, but some types were considerably under-researched. Online and web-based (12 studies) was the most researched category, although this covered a wide range of different products, and included weblogs or blogs, which might have been considered a separate category. The next most frequent was MOOCs (ten studies), although these were very various and suffered from very low completion rates and often even lower PA rates. Projects using videos were next in frequency (nine studies).

Five studies involved the use of Facebook as a platform for sharing and PA of videos or pictures. While Facebook and other social media platforms may be acceptable in higher education and beyond, there are issues about their use in a school context, where they are often forbidden by local authorities, so teachers would have inherent difficulties in using such platforms as a means of encouraging peer interaction (although school peers might well choose to use such applications outside of school of their own volition). Additionally, social media platforms are numerous and constantly increasing, so it would be difficult for programme developers to know which application to use, although perhaps it would not matter if different platforms were used by different participants.

The remaining categories were considerably smaller. Digital frameworks to organize/structure PA involved three studies, and although widely used, rarely seemed to target teachers. Only two papers investigated e-portfolios, although PA of e-portfolios seems relatively easy to operationalise. Just one paper investigated the use of PDAs, perhaps because such a project involves the cost of supplying PDAs, while other digital interventions could operate on a bring-your-own-device basis. One paper investigated the use of

iPads, although presumably any kind of tablet would have served the purpose, and the issue is not so much the nature of the device, but what is done with it. A final single study was of the utility of wikis (although in this case there was a rare example of the intervention being carried through into use with school children).

Additionally, other kinds of digital technology were not yet reported as being used. An obvious example would be the use of mobile phones, acknowledged in the wider literature as a relatively low-cost device (which many pre- and in-service teachers already own), and capable of supporting the viewing of videos (multiply if necessary) as well as Facebook and other social media.

Of the 37 secondary variables, probably the 15 most practically important are: formative or summative, compulsory vs. voluntary, pairs or groups, matching strategy, frequency, duration, anonymous or identified, training provided, participants involved in designing criteria, rubric used, additional prompts or scaffolds given, guidance on form of feedback given, complexity of work, monitoring or quality control, and alignment with or contribution to final formal assessment. Additionally, from a research point of view, attention to implementation integrity, follow-up, and the presence of ESs are highly desirable.

Most studies characterised their intervention as formative, and expected work to be improved as a result. Half the studies did not specify whether the PA was voluntary or compulsory. Of those that did, more said that it was compulsory than voluntary. This choice seemed to be determined by factors other than the strategic, since one might expect PA to start by being voluntary and progress to being accepted by all.

Grouping was usually in small groups (typically  $n = 4$ ), with all members assessing all other members, although a minority operated in reciprocal pairs, but some studies did not report the nature of grouping. The nature of matching was given even less thought. Although it was mostly same-age, and sometimes random allocation was used, generally the issue was not addressed. This raises interesting issues regarding effectiveness, since some groups would be very different in character from other groups. Where thought is given to this issue, matching is usually done to ensure a range of ability within each group. Likewise, the issue of consistency in each group membership over the course of the PA was generally not addressed, only six studies reporting consistency and one reporting deliberate change.

Regarding complexity of work, the studies reported here showed great variety in complexity. Almost a quarter of them required assessment of actual teaching, either to peers in a micro-teaching setting or to real school students in a classroom – clearly a highly complex activity. Another substantial section had more timorous and somewhat more theoretical objectives, involving developing a teaching plan or a research proposal. One would expect PA to start with the assessment of less complex tasks and build up to the assessment of more complex tasks.

About a quarter of studies reported on anonymity vs. identification of participants, and these were equally divided between the two. Some studies of the relative effectiveness of anonymity vs. identification are needed, but these need to be done over a period, since anonymity may be better in the short term but identification better in the long term, for instance.

Beyond this, the number and gender of the participants must be reported, although in one study even the number of participants was not reported, and in several studies the gender balance was not reported. Participant numbers do vary enormously, MOOCs

having very high numbers with low completion rates and many other studies quite modest numbers but with high completion rates. Pre-service teachers can be expected to be predominantly female in most countries, and there may be a difference in female to male responding to PA, but at the moment the picture is somewhat confused and more data is needed.

The subject studied must be reported, which over a quarter failed to do. While some studies were in science, others were in creativity or physical education. Clearly this is likely to affect the nature of PA, since in creativity there should be greater acknowledgement that the feedback given is highly subjective. Consequently, there might be more need for feedback from more than one assessor, and the assessee might be more selective in deciding what aspects of their work to alter.

Studies should specify clearly whether they have operated peer grades or elaborated peer feedback or both. In this review feedback alone was the most popular, closely followed by feedback with grades, but in 16% of studies even this was not reported. The research literature beyond teacher training says something about the relative effectiveness of these three methods, so it would be interesting and practically important to see if these findings also apply to teacher training.

The year of degree for pre-service teacher assessors and assessees should also be reported, which most studies neglected to do. In the West, for primary school pre-service teachers a three-year course is usual, while for secondary school teachers a one-year course is usual. Thus, while secondary teachers might only experience one type of PA in their course, primary teachers might experience several. Additionally, as primary teachers experience PA in successive years, they are likely to become more proficient in it. Further, there is the possibility of cross-year PA, with older and more experienced students acting as assessors for the less experienced, creating a form of apprenticeship.

The nature of the product or products assessed should also be specified. Assessing a presentation is a different matter from assessing a piece of writing, and somewhat different skills may apply. Generalisation from assessing one type of product to another type of product cannot be assumed. Participants may reasonably expect to start PA with relatively simple tasks and progress to more difficult tasks, but establishing a hierarchy of difficulty for types of product for this purpose is not straightforward.

The frequency, duration and length of the PA should always be reported, which in over a quarter of cases was not. How many times per week should PA occur, for what length of time on each occasion, and for how many weeks? The few studies here which reported this varied from one to eight rounds of PA, over four to nine weeks. These simple indicators of dosage are crucial in comparing studies.

The country of origin is of interest, as different countries are likely to have different cultural variables which influence the intervention – not necessarily its effectiveness, but the acceptability of different components and the way in which it is effective.

The nature of the teachers with whom the intervention is undertaken must be reported in studies, and generally was. Studies with pre-service teachers greatly outnumbered studies with in-service teachers, doubtless owing to the accessibility and likely compliance of pre-service teachers to researchers in universities, but this then created problems of follow-up (which may or may not have occurred to the researchers as desirable), although follow-up across years of a three-year degree was still entirely possible. Studies of pre-service teachers were over three times more prevalent than studies of in-service teachers, although in-service

teachers were in a better position to use their experiences to implement PA with their students in the classroom. Clearly this imbalance needs to be addressed.

Involving participants in designing assessment criteria enhances their engagement with PA, yet only two papers reported doing this. Providing training enhances outcomes from PA, but only 13 (30%) studies reported this, and in only four cases did the training include actual practice at PA. Studies need to involve participants in designing assessment criteria, and provide training which includes practice at PA. Half of the studies used a rubric to scaffold the PA, but only two studies reported deploying additional prompts or scaffolds, via hyperlinks or case studies. Projects need to involve participants in using assessment rubrics and make additional prompts or scaffolds available.

It was very unusual for the degree of alignment with any subsequent form of official assessment to be reported, although clearly this was likely to affect the participants' view of how worthwhile the PA had been.

The nature of the measures used should be clearly characterised in terms of reliability and validity, and studies should seek to use more than one kind of measure. Reliability and validity are particularly uncertain in relation to subjective participant perceptions of success or otherwise in PA. Some studies used this as the only outcome measure, which is clearly unsatisfactory. While participant perceptions may indeed have a useful role to play in investigations, they should not form the only means of gathering data in a study. Additionally, some studies used such measures on only one occasion, so it was impossible to grasp any change which might have occurred during the programme, and future studies need to consider some form of pre-post assessment of participant perceptions.

The quality of studies was generally low, and attempts to analyse the quality of studies was found to be too unreliable to permit its reporting (Guyatt et al. 2011). For future research, it may be that the criteria for inclusion/exclusion will need to be stricter, and quality outliers eliminated from the analysis. Although this might make quality categorisation possible, it is complex when both quantitative and qualitative studies are to be accepted. Nonetheless, eliminating studies which only report participant perceptions would be a step in this direction. A more stringent alternative would be to eliminate all studies which only report a single outcome measure, but this might leave the number of included studies too low.

Thus, future research should: use more objective measures than surveys or interviews, use more than one kind of outcome measure, use control or comparison groups if at all possible, give ESs, investigate implementation integrity and follow-up, and investigate subsequent PA use with school children. None of these studies made any mention of implementation fidelity/integrity or follow-up, and only two papers gave ESs. Comparing different types of PA of similar length in similar subject areas would be very valuable. The moderator variable analysis as above is likely to be useful to researchers for reviews and meta-analyses of PA in other areas, and indeed to practitioners in designing and reporting their own form of PA.

### ***6.5. Implications for practitioners and policymakers***

The moderator variable analysis used here is likely to be useful to researchers for reviews and meta-analyses of PA in other areas. Perhaps even more importantly, it should be useful to practitioners setting out to design (and hopefully report) their own form of PA. While PA in teacher training and development generally works, and teachers can be

reassured by this, its effectiveness could be increased by careful attention these planning variables. Beyond this, PA should be carried forward from teacher training and development into practical projects with school pupils in the classroom. Indeed, this would arguably be the most appropriate test for the effectiveness of the teachers' own PA experiences. The school projects would need reporting so that evidence is accumulated of any doubly additive effect of PA.

From a policymaker perspective, they should be aware that PA is a conglomerate term, and require clearer specification in relation to the moderator variables of the nature of PA intended when commissioning research or practice. They should also be encouraging teacher trainers and schools to be developing projects which start with PA for teachers and extend that into PA for school pupils, treating the whole cycle as an integrated project.

## 7. Conclusions

In this systematic review, digital PA in school teacher education/development had mainly positive effects, only one study out of 43 showing negative effects (RQ1 affirmative). No study showed awareness of the whole range of moderator variables and many gaps were identified (RQ2 negative). Given the quality of studies, caution in interpretation is needed. While PA can be confidently recommended for implementation in teacher education and development, practitioners and researchers need to give much more thought to *what kind* and *how* it is implemented. The moderator variable analysis presented here should not only give teachers useful ideas about implementation, but also help future researchers to ensure that all variables are fully described in research reports. Once we have more controlled studies as well as a full reporting of all moderator variables, we will be in a stronger position to move forward.

## Acknowledgments

Thanks are due to European Schoolnet who funded this research. Thanks also to Benjamin Hertz and Patricia Wastiau who commented constructively on an earlier draft of this paper.

## Disclosure statement

The results of this study do not create a conflict of interest for the author.

## Funding

This work was supported by the European Schoolnet, [www.eun.org](http://www.eun.org), no grant number

## Notes on contributor

*Keith James Topping* is Professor of Educational and Social Research in the School of Education at the University of Dundee. His research interests include peer learning, computer assisted learning and assessment, and inclusion, with over 425 research publications in 16 languages. Further details are at [https://en.wikipedia.org/wiki/Keith\\_James\\_Topping](https://en.wikipedia.org/wiki/Keith_James_Topping) and <https://www.dundee.ac.uk/esw/staff/details/toppingkeith-j-.php#tab-bio>.



## ORCID

Keith James Topping  <http://orcid.org/0000-0002-0589-6796>

## Data Statement

The systematically analysed papers are available from the author on request.

## References

Items in the systematic analysis are asterisked \*

Items in the systematic analysis are asterisked \*

- Alexander, P. A. 2020. "Methodological Guidance Paper: The Art and Science of Quality Systematic Reviews." *Review of Educational Research* 90 (1): 6–23. doi:10.3102/0034654319854352.
- \*Anat, K., K. Einav, and R. Shirley. 2020. "Development of Mathematics Trainee Teachers' Knowledge while Creating a MOOC." *International Journal of Mathematical Education in Science and Technology* 51 (6): 939–953. doi: 10.1080/0020739X.2019.1688402
- Annis, L. F. 1983. "The Processes and Effects of Peer Learning." *Human Learning* 2 (1): 39–47.
- \*Bachelet, R., D. Zongo, and A. Bourelle. 2015. "Does Peer Grading Work? How to Implement and Improve It? Comparing Instructor and Peer Assessment in MOOC GdP." Paper delivered at European MOOCs Stakeholders Summit 2015, May 2015, Mons, Belgium. <https://halshs.archives-ouvertes.fr/halshs-01146710v2>
- \*Backhouse, A., I. Wilson, and D. Mackley. 2015. "Using iPads to Increase the Level of Student Engagement in the Peer Review and Feedback Process." In *Ipads in Higher Education: Proceedings of the 1st International Conference*, edited by N. Souleles and C. Pillar. Newcastle: Cambridge Scholars Publishing. <http://eprints.lincoln.ac.uk/24654/1/24654%20iPads%20in%20HE%20ABackhouse.p>
- \*Bonk, C. J., K. Daytner, G. Daytner, V. Dennen, and S. Malikowski. 2001. "Using Web-based Cases to Enhance, Extend, and Transform Preservice Teacher Training: Two Years in Review." *Computers in the Schools* 18 (1): 189–211. doi: 10.1300/J025v18n01\_01
- \*Borowczak, M., and A. C. Burrows. 2016. "Enabling Collaboration and Video Assessment: Exposing Trends in Science Preservice Teachers' Assessments." *Contemporary Issues in Technology and Teacher Education (CITE Journal)* 16 (2): 127–150. <https://www.learntechlib.org/primary/p/161911>
- Bozkurt, A., E. Akgün-Özbek, and O. Zawacki-Richter. 2017. "Trends and Patterns in Massive Open Online Courses: Review and Content Analysis of Research on MOOCs (2008–2015)." *The International Review of Research in Open and Distributed Learning*, 18(5). <http://www.irrodl.org/index.php/irrodl/article/download/3080/4284?inline=1.MOO>
- Brooks, C., A. Carroll, R. M. Gillies, and J. Hattie. 2019. "A Matrix of Feedback for Learning." *Australian Journal of Teacher Education* 44 (4): 14–32. doi:10.14221/ajte.2018v44n4.2.
- \*Çelik, S., E. Baran, and O. Sert. 2018. "The Affordances of Mobile-App Supported Teacher Observations for Peer Feedback." *International Journal of Mobile and Blended Learning*, 10 (2): 36–49. doi: 10.4018/IJMBL.2018040104
- \*Chang, C. C. 2002. "Assessing and Analyzing the Effects of WBLP on Learning Processes and Achievements: Using the Electronic Portfolio for Authentic Assessment on University Students' Learning." In *Proceedings of ED-MEDIA 2002 - World Conference on Educational Multimedia, Hypermedia & Telecommunications*, edited by P. Barker and S. Rebelsky, 265–270. Denver, CO: Association for the Advancement of Computing in Education (AACE). <https://www.learntechlib.org/primary/p/9865>
- \*Chen, C. H. 2010. "The Implementation and Evaluation of a Mobile Self- and Peer-Assessment System." *Computers & Education* 55 (1): 229–236. doi:10.1016/j.compedu.2010.01.008.

- \*Chen, Y. C., and C. C. Tsai. 2009. "An Educational Research Course Facilitated by Online Peer Assessment." *Innovations in Education and Teaching International* 46 (1): 105–117. doi: [10.1080/14703290802646297](https://doi.org/10.1080/14703290802646297)
- Cheng, K. H., and C. C. Tsai. 2012. "Students' Interpersonal Perspectives On, Conceptions of and Approaches to Learning in Online Peer Assessment." *Australasian Journal of Educational Technology* 28 (4). doi:[10.14742/ajet.830](https://doi.org/10.14742/ajet.830).
- Darling-Hammond, L. 2010. "Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching". *ERIC Number: ED535859*. Washington, DC: Center for American Progress. <https://files.eric.ed.gov/fulltext/ED535859.pdf>
- Darling-Hammond, L., S. P. Newton, and R. C. Wei. 2013. "Developing and Assessing Beginning Teacher Effectiveness: The Potential of Performance Assessments". *Educational Assessment, Evaluation and Accountability* 25: 179–204. doi:[10.1007/s11092-013-9163-0](https://doi.org/10.1007/s11092-013-9163-0).
- \*De Marsico, M., F. Sciarroni, A. Sterbini, and M. Temperini. 2018. "Peer Assessment and Knowledge Discovering in a Community of Learners." Paper given at 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Seville, Spain. doi: [10.5220/0007229401190126](https://doi.org/10.5220/0007229401190126).
- \*Demir, M. 2018. "Using Online Peer Assessment in an Instructional Technology and Material Design Course through Social Media." *Higher Education: The International Journal of Higher Education Research* 75 (3): 399–414. doi:[10.1007/s10734-017-0146-9](https://doi.org/10.1007/s10734-017-0146-9).
- Deneen, C. C., G. W. Fulmer, G. T. L. Brown, K. Tan, W. S. Leong, and H. Y. Tay. 2019. "Value, Practice and Proficiency: Teachers' Complex Relationship with Assessment for Learning." *Teaching and Teacher Education*, 80: 39–47. doi:[10.1016/j.tate.2018.12.0220742-051X](https://doi.org/10.1016/j.tate.2018.12.0220742-051X).
- Double, K. S., J. A. McGrane, and T. N. Hopfenbeck. 2020. "The Impact of Peer Assessment on Academic Performance: A Meta-Analysis of Control Group Studies." *Educational Psychology Review*, 32: 481–509. doi: [10.1007/s10648-019-09510-3](https://doi.org/10.1007/s10648-019-09510-3)
- EPPI-Centre. 2004. "A Systematic Review of the Evidence of Reliability and Validity of Assessment by Teachers Used for Summative Purposes". London: Institute of Education, University of London. [http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/ass\\_rv3.pdf?ver=2006-03-02-124720-170](http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/ass_rv3.pdf?ver=2006-03-02-124720-170)
- \*Ersöz, Y., and S. N. Sad. 2018. "Facebook as A Peer-Assessment Platform: A Case Study in Art Teacher Education Context." *International Journal of Assessment Tools in Education* 5 (4): 740–753. doi: [10.21449/ijate.478277](https://doi.org/10.21449/ijate.478277).
- Falchikov, N., and J. Goldfinch. 2000. "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks." *Review of Educational Research* 70 (3): 287–322. doi:[10.3102/00346543070003287](https://doi.org/10.3102/00346543070003287).
- \*Foschi, L. C., and G. Cecchinato. 2019. "Validity and Reliability of Peer-Grading in In-Service Teacher Training." *Italian Journal of Educational Research* 177–194. <https://80.211.104.80/index.php/sird/article/view/3276>
- Friedman, B. A., P. L. Cox, and L. Maher. 2008. "An Expectancy Theory Motivation Approach to Peer Assessment." *Journal of Management Education*, 32: 580–612. doi:[10.1177/1052562907310641](https://doi.org/10.1177/1052562907310641)
- Fu, Q. K., C. J. Lin, and G. J. Hwang. 2019. "Research Trends and Applications of Technology-Supported Peer Assessment: A Review of Selected Journal Publications from 2007 to 2016." *Journal of Computers in Education* 6 (2): 191–213. doi:[10.1007/s40692-019-00131-x](https://doi.org/10.1007/s40692-019-00131-x).
- \*Gamage, D., M. E. Whiting, T. Rajapakshe, H. Thilakarathne, P. Indika, and S. Fernando. 2017. "Improving Assessment on Moocs through Peer Identification and Aligned Incentives." *Proceedings of the Fourth (2017) ACM Conference on Learning@Scale*. doi: [10.1145/3051457.3054013](https://doi.org/10.1145/3051457.3054013)
- Gielen, S., F. Dochy, and P. Onghena. 2011. "An Inventory of Peer Assessment Diversity." *Assessment & Evaluation in Higher Education* 36 (2): 137–155. doi:[10.1080/02602930903221444](https://doi.org/10.1080/02602930903221444).
- \*Goeze, A., J. M. Zottmann, F. Vogel, F. Fischer, and J. Schrader. 2014. "Getting Immersed in Teacher and Student Perspectives? Facilitating Analytical Competence Using Video Cases in Teacher Education." *Instructional Science* 42 (1): 91–114. doi [10.1007/s11251-013-9304-3](https://doi.org/10.1007/s11251-013-9304-3)

- \*Gogoulou, A., E. Gouli, M. Grigoriadou, M. Samarakou, and D. Chinou. 2007. "A Web-Based Educational Setting Supporting Individualized Learning, Collaborative Learning and Assessment." *Educational Technology and Society* 10 (4): 242–256. <https://www.jstor.org/stable/jeductechsoci.10.4.242>
- Guyatt, G., A. D. Oxman, E. A. Akl, R. Kunz, G. Vist, J. Brozek, S. Norris, et al. 2011. "GRADE Guidelines: 1. Introduction - GRADE Evidence Profiles and Summary of Findings Tables." *Journal of Clinical Epidemiology* 64 (4): 383–394. doi:10.1016/j.jclinepi.2010.04.026.
- Harlen, W. 2005. "Trusting Teachers' Judgement: Research Evidence of the Reliability and Validity of Teachers' Assessment Used for Summative Purposes." *Research Papers in Education* 20 (3): 245–270. doi:10.1080/02671520500193744.
- Hou, H. T., T. F. Yu, F. D. Chiang, Y. H. Lin, K. E. Chang, and C. C. Kuo. 2020. "Development and Evaluation of Mindtool-Based Blogs to Promote Learners' Higher Order Cognitive Thinking in Online Discussions: An Analysis of Learning Effects and Cognitive Process." *Journal of Educational Computing Research* 58 (2): 343–363. doi:10.1177/0735633119830735.
- Jacoby, J. 2014. "The Disruptive Potential of the Massive Open Online Course: A Literature Review." *Journal of Open, Flexible and Distance Learning* 18 (1): 73. <https://www.learntechlib.org/p/148551>
- Johnson, S. 2013. "On the Reliability of High-Stakes Teacher Assessment." *Research Papers in Education* 28 (1): 91–105. doi:10.1080/02671522.2012.754229.
- Kennedy, J. 2014. "Characteristics of Massive Open Online Courses (Moocs): A Research Review, 2009-2012." *Journal of Interactive Online Learning* 13 (1): 1–16.
- \*Lamb, P., K. Lane, and D. Aldous. 2012. "Enhancing the Spaces of Reflection: A Buddy Peer-Review Process within Physical Education Initial Teacher Education." *European Physical Education Review* 19 (1): 21–38.
- \*Laurillard, D. 2016. "The Educational Problem that MOOCs Could Solve: Professional Development for Teachers of Disadvantaged Students." *Research in Learning Technology* 24 (1): 29369. doi:10.3402/rlt.v24.29369.
- Li, H., Y. Xiong, C. V. Hunter, X. Guo, and R. Tywniwi. 2020. "Does Peer Assessment Promote Student Learning? A Meta-Analysis." *Assessment and Evaluation in Higher Education* 45 (2): 193–211. doi:10.1080/02602938.2019.1620679.
- Li, L. 2017. "The Role of Anonymity in Peer Assessment." *Assessment and Evaluation in Higher Education* 42 (4): 645–656, doi: 10.1080/02602938.2016.1174766.
- Li, L., X. Y. Liu, and Y. C. Zhou 2012. "Give and Take: A Re-Analysis of Assessor and Assessee's Roles in Technology-Facilitated Peer Assessment." *British Journal of Educational Technology* 43 (3): 376–384. doi:10.1111/j.1467-8535.2011.01180.x.
- \*Lin, G. Y. 2016. "Effects that Facebook-Based Online Peer Assessment with Micro-Teaching Videos Can Have on Attitudes toward Peer Assessment and Perceived Learning from Peer Assessment." *EURASIA Journal of Mathematics, Science & Technology Education* 12 (9): 2295–2307. doi:10.12973/eurasia.2016.1280a.
- \*Liu, X. Y., and L. Li. 2014. "Assessment Training Effects on Student Assessment Skills and Task Performance in a Technology-Facilitated Peer Assessment." *Assessment & Evaluation in Higher Education* 39 (3): 275–292, doi: 10.1080/02602938.2013.823540.
- \*Luo, H., A. C. Robinson, and J. Y. Park. 2014. Peer Grading in a MOOC: Reliability, Validity, and Perceived Effects. *Online Learning Journal* 18 (2): 1–14. <https://www.learntechlib.org/p/183756>
- Luxton-Reilly, A. 2009. "A Systematic Review of Tools that Support Peer Assessment." *Computer Science Education* 19 (4): 209–232. doi: 10.1080/08993400903384844.
- McLuckie, J., and K. Topping. (2004). "Transferable Skills for Online Peer Learning." *Assessment & Evaluation in Higher Education* 29 (5): 563–584. doi: 10.1080/02602930410001689144
- \*Mohamed, M. H., and M. Hammond. 2018. "Moocs: A Differentiation by Pedagogy, Content and Assessment." *International Journal of Information and Learning Technology* 35 (1): 2–11. doi:10.1108/IJILT-07-2017-0062.
- Moher, D., A. Liberati, J. Tetzlaff, and D. G. Altman. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *PLoS Medicine* 6 (7): e1000097. doi:10.1371/journal.pmed.1000097.

- \*Ng, E. M. W. 2016. "Fostering Pre-Service Teachers' Self-Regulated Learning through Self- and Peer Assessment of Wiki Projects." *Computers & Education* 98: 180–191. doi:[10.1016/j.compedu.2016.03.015](https://doi.org/10.1016/j.compedu.2016.03.015).
- O'Donnell, A. M., and K. J. Topping. 1998. "Peers Assessing Peers: Possibilities and Problems." In *Peer-Assisted Learning*, edited by K. Topping and S. Ehly. Mahwah, NJ: Lawrence Erlbaum.
- \*Okumuş, K., and L. H. Yurdakal. 2016. "Peer Feedback through SNSs (Social Networking Sites): Student Teachers' Views about Using Facebook for Peer Feedback on Microteachings." *Elementary Education Online* 15 (4): 1206–1216. doi: [10.17051/eeo.2016.17666](https://doi.org/10.17051/eeo.2016.17666)
- Orwin, R. G. 1983. "A Fail-Safe N for Effect Size in Meta-Analysis." *Journal of Educational and Behavioral Statistics* 8 (2): 157–159. doi:[10.3102/10769986008002157](https://doi.org/10.3102/10769986008002157).
- Parsons, S. A., A. C. Hutchison, L. A. Hall, A. W. Parsons, S. T. Ives, and A. B. Leggett. 2019. "U.S. Teachers' Perceptions of Online Professional Development." *Teaching and Teacher Education: An International Journal of Research and Studies* 82 (1): 33–42. <https://www.learntechlib.org/p/208294>
- Pecheone, R. L., and R. R. Chung. 2006. "Evidence in Teacher Education: The Performance Assessment for California Teachers (PACT)." *Journal of Teacher Education* 57 (1): 22–36. doi:[10.1177/0022487105284045](https://doi.org/10.1177/0022487105284045).
- \*Picci, P., A. Calvani, and G. Bonaiuti. 2012. The Use of Digital Video Annotation in Teacher Training: The Teachers' Perspectives." *Procedia - Social and Behavioral Sciences* 69: 600–613. doi: [10.1016/j.sbspro.2012.11.452](https://doi.org/10.1016/j.sbspro.2012.11.452).
- \*Pöldoja, H., T. Väljataga, M. Laanpere, and K. Tammets. 2012. "Web-based Self- and Peer-Assessment of Teachers' Digital Competencies." *World Wide Web* 17 (2): 255–269. doi: [10.1007/s11280-012-0176-2](https://doi.org/10.1007/s11280-012-0176-2).
- \*Ramdani, J. M., and H. P. Widodo. 2019. "Student Teachers' Engagement in Facebook-Assisted Peer Assessment in an Initial Teacher Education Context: Speaking 2.0." *Journal of Education for Teaching: International Research and Pedagogy* 45 (3): 348–352. doi: [10.1080/09589236.2019.1599503](https://doi.org/10.1080/09589236.2019.1599503)
- \*Savas, P. 2012. Micro-teaching Videos in EFL Teacher Education Methodology Courses: Tools to Enhance English Proficiency and Teaching Skills among Trainees." *Procedia - Social and Behavioral Sciences* 55: 730–738. doi: [10.1016/j.sbspro.2012.09.558](https://doi.org/10.1016/j.sbspro.2012.09.558)
- \*Seifert, T., and O. Feliks. 2019. "Online Self-Assessment and Peer-Assessment as a Tool to Enhance Student-Teachers' Assessment Skills." *Assessment & Evaluation in Higher Education* 44 (2): 169–185. doi: [10.1080/02602938.2018.1487023](https://doi.org/10.1080/02602938.2018.1487023)
- \*Sert, O., and A. Aşık. 2020. "A Corpus Linguistic Investigation into Online Peer Feedback Practices in CALL Teacher Education." *Applied Linguistics Review* 11 (1): 55–78. doi:[10.1515/applirev-2017-0054](https://doi.org/10.1515/applirev-2017-0054).
- \*So, W. M., H. K. Hung, and Y. W. Yip. 2008. "The Digital Video Database: A Virtual Learning Community for Teacher Education." *Australasian Journal of Educational Technology* 24 (1): 73–90. doi:[10.14742/ajet.1231](https://doi.org/10.14742/ajet.1231).
- \*Sterbini, A., and M. Temperini. 2013. "Peer-assessment and Grading of Open Answers in a Web-Based E-Learning Setting." *2013 12th International Conference on Information Technology Based Higher Education and Training (ITHET)*, October, 1–7. doi: [10.1109/ITHET.2013.6671056](https://doi.org/10.1109/ITHET.2013.6671056)
- Tenorio, T., I. I. Bittencourt, S. Isotani, and A. P. Silva. 2016. "Does Peer Assessment in On-Line Learning Environments Work? A Systematic Review of the Literature." *Computers in Human Behavior* 64: 94–107. doi:[10.1016/j.chb.2016.06.020](https://doi.org/10.1016/j.chb.2016.06.020).
- Topping, K. J. 1998. "Peer Assessment between Students in College and University." *Review of Educational Research* 68 (3): 249–276. doi:[10.3102/00346543068003249](https://doi.org/10.3102/00346543068003249).
- Topping, K. J. 2018. "Student Assessment for Educators Series." In *Using Peer Assessment to Inspire Reflection and Learning*, edited by J. H. MacMillan. New York & London: Routledge. ISBN: 978-0-8153-6765-9 (pbk). [www.routledge.com/9780815367659](http://www.routledge.com/9780815367659)
- \*Tsai, C. C. 2012. "The Development of Epistemic Relativism versus Social Relativism via Online Peer Assessment, and Their Relations with Epistemological Beliefs and Internet Self-Efficacy."

- Educational Technology & Society* 15 (2): 309–316. <https://www.jstor.org/stable/jeductechsoci.15.2.309>
- \*Tsai, C. C., and J. C. Liang. 2009. “The Development of Science Activities via On-Line Peer Assessment: The Role of Scientific Epistemological Views.” *Instructional Science: An International Journal of the Learning Sciences* 37 (3): 293–310. doi: [10.1007/s11251-007-9047-0](https://doi.org/10.1007/s11251-007-9047-0)
- \*Tsai, C. C., S. S. J. Lin, and S. M. Yuan. 2002. “Developing Science Activities through a Networked Peer Assessment System.” *Computers & Education* 38 (1–3): 241–252. doi:[10.1016/S0360-1315\(01\)00069-0](https://doi.org/10.1016/S0360-1315(01)00069-0).
- \*Tsai, C. C., E. Z. F. Liu, S. S. J. Lin, and S. M. Yuan. 2001. “A Networked Peer Assessment System Based on A Vee Heuristic.” *Innovations in Education and Teaching International* 38 (3): 220–230. doi: [10.1080/14703290110051415](https://doi.org/10.1080/14703290110051415)
- \*Tsivitanidou, O. E., and C. P. Constantinou. 2016. “A Study of Students’ Heuristics and Strategy Patterns in Web-Based Reciprocal Peer Assessment for Science Learning.” *The Internet and Higher Education*, 29: 12–22. doi:[10.1016/j.iheduc.2015.11.002](https://doi.org/10.1016/j.iheduc.2015.11.002).
- Van Popta, E., M. Kral, G. Camp, R. L. Martens, and P. R. Simons. 2017. “Exploring the Value of Peer Feedback in Online Learning for the Provider.” *Educational Research Review* 20: 24–34. doi:[10.1016/j.edurev.2016.10.003](https://doi.org/10.1016/j.edurev.2016.10.003).
- Veletsianos, G., and P. Shepherdson. 2016. “A Systematic Analysis and Synthesis of the Empirical MOOC Literature Published in 2013–2015.” *The International Review of Research in Open and Distributed Learning* 17 (2): 198–221. doi: [10.19173/irrodl.v17i2.2448](https://doi.org/10.19173/irrodl.v17i2.2448)
- \*Vivian, R., K. Falkner, and N. Falkner. 2014. Addressing the Challenges of a New Digital Technologies Curriculum: MOOCs as a Scalable Solution for Teacher Professional Development.” *Research in Learning Technology* 22: 1–19. doi:[10.3402/rlt.v22.24691](https://doi.org/10.3402/rlt.v22.24691).
- Vu, L. T. 2017. “A Case Study of Peer Assessment in A MOOC-based Composition Course: Students’ Perceptions, Peers’ Grading Scores versus Instructors’ Grading Scores, and Peers’ Commentary versus Instructors’ Commentary.” Doctor of Philosophy Degree Dissertation, Southern Illinois University Carbondale. <https://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?referer=http://www.jurn.org/&httpsredir=1&article=2398&context=dissertations>
- \*Welsh, M. 2012. “Student Perceptions of Using the Pebblepad E-Portfolio System to Support Self- and Peer-Based Formative Assessment.” *Technology, Pedagogy and Education* 21 (1): 57–83. doi: [10.1080/1475939X.2012.659884](https://doi.org/10.1080/1475939X.2012.659884)
- \*Wen, M. C. L., and C. C. Tsai. 2008. “Online Peer Assessment in an Inservice Science and Mathematics Teacher Education Course.” *Teaching in Higher Education* 13 (1): 55–67. doi: [10.1080/13562510701794050](https://doi.org/10.1080/13562510701794050)
- \*Wopereis, I. G. J. H., P. B. Sloep, and S. H. Poortman. 2010. “Weblogs as Instruments for Reflection on Action in Teacher Education.” *Interactive Learning Environments* 18 (3): 245–261. doi: [10.1080/10494820.2010.500530](https://doi.org/10.1080/10494820.2010.500530)
- \*Wu, C. C., and H. C. Kao. 2008. “Streaming Videos in Peer Assessment to Support Training Pre-Service Teachers.” *Educational Technology & Society* 11 (1): 45–55. <https://www.jstor.org/stable/jeductechsoci.11.1.45>
- Yim, S., and Y. Cho. 2016. Predicting Pre-Service Teachers’ Intention of Implementing Peer Assessment for Low-Achieving Students *Asia Pacific Education Review* 17 (1): 63–72. doi: [10.1007/s12564-016-9416-y](https://doi.org/10.1007/s12564-016-9416-y)
- Yu, F. Y. 2011. “Multiple Peer-Assessment Modes to Augment Online Student Question-Generation Processes.” *Computers & Education* 56 (2): 484–494. doi:[10.1016/j.compedu.2010.08.025](https://doi.org/10.1016/j.compedu.2010.08.025).
- Yu, S. L., and I. Lee. 2016. “Peer Feedback in Second Language Writing (2005–2014).” *Language Teaching* 49 (4): 461–493. doi:[10.1017/S0261444816000161](https://doi.org/10.1017/S0261444816000161)
- Zheng, L. Q., X. Zhang, and P. P. Cui. 2020. “The Role of Technology-Facilitated Peer Assessment and Supporting Strategies: A Meta-Analysis.” *Assessment & Evaluation in Higher Education* 45 (3): 372–386. doi: [10.1080/02602938.2019.1644603](https://doi.org/10.1080/02602938.2019.1644603).
- Zhou, J., Y. Zheng, and J. Tai. 2020. “Grudges and Gratitude: The Social-Affective Impacts of Peer Assessment.” *Assessment & Evaluation in Higher Education* 45 (3): 345–358. doi:[10.1080/02602938.2019.1643449](https://doi.org/10.1080/02602938.2019.1643449)